

J. P. DESCLES

**Programme interdisciplinaire de traitement formel et automatique
des langues et du langage (PITFALL)**

Mathématiques et sciences humaines, tome 77 (1982), p. 43-91

http://www.numdam.org/item?id=MSH_1982__77__43_0

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1982, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

PROGRAMME INTERDISCIPLINAIRE DE TRAITEMENT FORMEL ET AUTOMATIQUE DES
DES LANGUES ET DU LANGAGE (PITFALL)

J.P. DESCLES*

AVERTISSEMENT

L'article qui suit présente les hypothèses initiales et la méthodologie d'un projet né en 1974 lors d'un contrat DGRST (Informatique et Sciences Humaines). Ce dernier avait pour objectif de donner une formulation plus technique au cadre programmatique proposé par A. Culioli dans son article de 1968 paru dans les Cahiers pour l'Analyse.

Ce programme est poursuivi par une équipe interdisciplinaire de recherche (Equipe de Recherche Associée 642 du CNRS) il est soutenu par diverses Actions Thématiques Programmées du CNRS (Informatique et Sciences Humaines, Intelligence Artificielle).

La mise en place d'une théorie des invariants langagiers, objet principal de ce programme de travail, est encore en pleine gestation. Peut-il en être autrement ? Cette théorie n'a pas encore produit une publication générale - mis à part le rapport présenté à l'UNESCO en 1979 - qui exposerait de façon relativement succincte et conceptuelle la liste des principes retenus et la méthode utilisée. Ce texte est une reprise entièrement refondue d'un rapport scientifique écrit en 1976. Il vise à combler provisoirement cette lacune, mais, comme on sait, la formulation ultime des hypothèses de travail ne peut être verbalisée qu'au terme d'une recherche et non à son origine.

Cette théorie des invariants repose sur une appréhension du langage et des langues que peut-être tous les "linguistes" contemporains ne partagent pas. Elle ne découle pas, en tout cas, de l'image que pourrait se former un non-spécialiste qui lirait les manuels de linguistique générale actuels. Pourtant, les problèmes abordés (langage et langues, relativité linguistique, construction des catégories grammaticales ...) font partie des préoccupations des linguistes, non seulement les classiques (Vendryès, Kuryowicz, Cohen, Sapir, Benveniste, Guillaume, Hjelmslev, les linguistes pragois, Jakobson pour ne citer que ceux-là) mais aussi les anciens médiévaux (avec l'Ecole des Terministes) et aussi les stoiciens. De plus, certaines précautions apparemment subtiles sont déjà présentes, au moins de façon latente, dans les travaux d'auteurs universels comme Frege ou

* Université Paris VII - UER de Mathématiques et Département de Recherches linguistiques, 2, Place Jussieu. 75221 Paris Cédex 05. (ERA 642 du CNRS)

Pierce qu'il est toujours bon de méditer.

Bien que ce programme ne propose pas un discours qui viendrait amplifier les vagues successives des "découvertes à la mode", il n'ignore pas pour autant la teneur conceptuelle et aussi les techniques d'analyse. Non seulement des théories chomskyennes mais aussi la théorie récente des opérateurs prédicatifs et transformationnels de Z.S. Harris ou encore la grammaire applicationnelle de S.K. Shaumyan, le modèle "sens-texte" de I.A. Mel'chouck, la grammaire universelle "à la Montague", les réseaux sémantiques, les grammaires relationnelles et les analyses par "Augmented Transition Network" utilisées en Intelligence Artificielle. Il n'est cependant pas question d'emprunter sans passage préalable au crible d'un doute systématique des distinctions "qui doivent être utiles puisque tout le monde ou presque les emploient".

Certes, la théorie serait jugée d'emblée "d'avant-garde" si elle reprenait des concepts trop employés et usés pour qu'ils aient conservé la force hypothétique des premiers écrits où ils avaient émergé, mais, que ce soient : signifié/signifiant, langue/parole, structure profonde/structure de surface, compétence/performance, syntaxe/sémantique/pragmatique, énoncés performatifs/énoncés constatifs, "compréhension" du langage naturel/analyse des langues etc. de telles oppositions adoptées sans examen obscurcissent souvent les débats en cachant et donc en oubliant certains des problèmes les plus fondamentaux et pourtant récurrents du langage et des langues. Celles-ci doivent être replacées dans un projet théorique. Il n'est pas possible non plus d'ignorer l'apport essentiel des théories contemporaines ne serait-ce que pour les problèmes qui ont été exhibés comme :

- . engendrement et reconnaissance par grammaires formelles et automates,*
- . recherche systématique et analyse des relations paraphrastiques,*
- . distinction entre un génotype invariant et abstrait et des phénotypes manifestés,*
- . prise en compte des phénomènes d'énonciation et ceux de nature foncièrement pragmatique,*
- . acquisition des structures et des mécanismes linguistiques par processus cognitifs,*
- . inférence naturelle*

ou pour les outils et modes de représentations qui en résultent :

- . logique combinatoire et λ -calcul,*
- . systèmes de réécriture et séries formelles,*
- . théorie des "modèles".*

L'objectif théorique montre bien qu'il ne s'agit pas d'une recherche isolée qui fabriquerait les seuls problèmes qu'elle serait capable de résoudre. Tout en repoussant la tentation de suivre la mode, le programme présenté ici veut

répondre en partie aux problèmes "classiques" de la linguistique et les présenter sous une forme qui tiendrait compte des techniques et préoccupations actuelles.

Le lecteur remarquera vite une certaine redondance, notamment au paragraphe 3 (Hypothèses fondamentales). La formulation claire et autonome des hypothèses de travail est indispensable mais une trop grande concision risquerait de paraître trop elliptique à tout lecteur pressé qui pourrait ainsi prétexter que le programme proposé est mal fondé et donc ne pourrait être validé.

La présentation qui suit est, répétons-le, conceptuelle. Dans un cadre plus général, par exemple celui d'un ouvrage, elle devrait être accompagnée d'exemples de linguistique empruntés à diverses langues, de rappels historiques sur le développement des disciplines logique, linguistique, mathématique, informatique. Tout ceci viendrait appuyer ces propos qui alors seraient mieux argumentés.

INTRODUCTION

L'approche interdisciplinaire est souvent très critiquée ; il s'agit souvent, pense-t-on, de substituer un discours hétéroclite à des discours cohérents. Nous indiquons très brièvement comment nous envisageons, pour notre part, l'approche interdisciplinaire qui doit tenir des discours cohérents dans chacune des disciplines concernées et non un seul discours incompréhensible aux uns et aux autres. Il s'agit d'indiquer également pourquoi un "traitement formel et automatique" est un grand secours, sinon une nécessité, pour qui recherche des invariants langagiers par rapport à la diversité des manifestations linguistiques, des énonciateurs et des langues. Les systèmes de représentations métalinguistiques sont externes aux langues naturelles et destinés à les décrire. Ces systèmes sont conçus comme des ensembles d'opérations (de production et de compréhension). Qui effectue les opérations ? Sur quels domaines portent les opérations ? Peut-on (et comment) définir une catégorie grammaticale ? Comment relier un système métalinguistique (composé d'opérations) à un système linguistique classificatoire ? Nous esquisserons simplement ces problèmes qui demanderaient de très longs développements techniques, méthodologiques et épistémologiques.

Comment concevoir un processus de validation d'hypothèses linguistiques? Quelles en sont les étapes? Y a-t-il une hiérarchie? Pourquoi construire des systèmes formels? On sait que toute expérimentation impose un certain nombre de précautions : quelles sont les données observables de départ? Comment enrichir ces données? Quelles sont les régularités structurelles qui apparaissent? Quels sont les choix explicites qui sont faits dans les données? Quelles sont les structures utilisées et pourquoi?

Les quelques pages qui suivent mentionnent donc ces problèmes sans pouvoir approfondir les questions soulevées car nous sommes conscients que c'est toute l'approche scientifique du langage et des langues qui est en jeu et qu'une discussion complète nécessiterait des centaines de pages.

1. VARIATIONS SUR UN TITRE

1.1. Interdisciplinarité et consensus

Le mot "interdisciplinaire" envahit notre univers scientifique quotidien: il est de bon ton d'"être interdisciplinaire", de "faire du pluridisciplinaire". La mode s'en est emparée. Pourtant, l'abus et la mauvaise utilisation de ce terme - employé parfois pour masquer une certaine incapacité - agace un grand nombre de chercheurs consciencieux. Cette notion paraît autoriser -sauf si on y prend bien garde- le recours à un imbroglio théorique; le mélange, sans précautions, des méthodes d'analyse; l'importation, sans interrogations préalables, de concepts d'un champ théorique dans un autre; des discours hétéroclites qui -du moins dans leur globalité- restent incompréhensibles aux uns et aux autres et donc ne se soumettent pas aux contrôles nécessaires et aux conditions de claire transmissibilité, deux caractéristiques indispensables à toute activité scientifique. En fait, une suspicion risque de toujours planer sur une recherche interdisciplinaire, à savoir l'incapacité de recevoir le consensus de la communauté scientifique. Or, si le consensus, plus ou moins large, de la "cité scientifique" est un des facteurs essentiels de

la recherche, il est également ressenti par chacun des chercheurs comme un gage indispensable de scientificité. Le consensus peut, bien entendu, ne pas être immédiate mais il est cependant révéélé, dans notre cité scientifique contemporaine, par de multiples voies: commissions scientifiques des organismes nationaux et internationaux, acceptations des communications ou invitations aux congrès et séminaires, publications d'articles ou de notes par des revues spécialisées, critiques organisées d'articles et de livres, soutenance de thèses, collations de diplômes... L'activité scientifique est ainsi soumise à un contrôle qui s'exerce (plus ou moins mal ou bien) dans le cadre de la communauté (inter-)nationale des scientifiques spécialistes du domaine où s'effectue la recherche évaluée, avec certes diverses pressions économiques, politiques, idéologiques, de mode et de tendances qui perturbent plus ou moins, selon les époques et les lieux, ce jeu des contrôles. La méthode d'approche interdisciplinaire risque alors, si on y réfléchit, de rendre caduques les conditions traditionnelles du contrôle par consensus: un travail interdisciplinaire qui ne peut être jugé dans sa globalité risque de ne plus entraîner une adhésion -même provisoire- qui serait obtenue à la suite d'une argumentation scientifique (réfutable éventuellement) dont chacun serait censé en connaître -au moins implicitement- les règles. Or, si les moyens de contrôle tendent à disparaître, la scientificité risque aussitôt de s'évanouir; personne ne peut et ne doit se porter garant de sa valeur car il n'est guère possible, à de très rares exceptions près, d'être à la fois juge, productif et créateur -à un haut niveau théorique et technique- dans plusieurs domaines disjoints qui ont chacun développé des techniques, des méthodes d'analyse et des concepts appropriés. Le consensus est d'autant plus difficile à réaliser que la méthode d'approche interdisciplinaire d'un même problème est souvent plus féconde lorsqu'elle articule des domaines a priori disjoints (médecine/mathématique; archéologie/informatique; droit/logique; linguistique/biologie; économie/physique...) et

plus normale, voire naturelle, lorsqu'elle fait intersecter des domaines proches ou de même nature (économie / science politique; grammaires française et anglaise; physique / chimie; chimie / biologie; linguistique de terrain / ethnologie...).

S'il y a un risque évident de perte de contrôle, l'interdisciplinarité bien comprise devrait chercher à y remédier par une rigueur accrue dans chacun des domaines concernés et par une réflexion poussée sur l'articulation théorique et méthodologique des disciplines mises en présence. De plus, et c'est là une option que nous prenons, il ne s'agit pas de produire un discours global, totalisant, souvent hétérogène et plus ou moins interprétable, mais de tenir des discours argumentés et distincts exprimés dans les langages de chacune discipline et donc compréhensibles et évaluables par chaque spécialiste des domaines concernés. Il ne s'agit pas non plus de juxtaposer des domaines mais, au contraire, d'articuler à la fois dans la construction théorique et dans les réalisations expérimentales, des domaines ayant en commun : (a) des problématiques compatibles sur un même thème; (b) une mise en place d'un champ conceptuel théorique qui nécessite une recherche approfondie dans chacun des domaines; (c) des techniques d'analyse expérimentale qui exigent des compétences complémentaires (à un haut niveau de qualité).

Le thème qui nous intéresse dans ce programme, c'est le langage et les langues (nous préciserons ce point plus loin); les domaines concernés sont la linguistique bien entendu et la mathématique et l'informatique (fondamentale et pratique). Ces domaines sont articulés et organisés entre-eux : la linguistique est première et doit veiller à ne pas transformer radicalement et sans justifications l'objet de son étude sous la pression exagérée des formalismes mathématiques et des techniques informatiques.

1.2. Traitement formel et automatique

Si la linguistique aspire à devenir une science, il lui faut tenir des discours cohérents, argumentés et réfutables; les hypothèses linguistiques sont alors obligatoirement formulées avec précision et soumises à vérification à l'aide d'un montage expérimental qui doit confronter le système des hypothèses aux phénomènes susceptibles d'être observés. Non seulement le linguiste se doit d'affiner ses procédures de raisonnement mais aussi il vise la plupart du temps à substituer aux manipulations métalinguistiques -plus ou moins transmissibles et univoques- de véritables procédures algorithmiques exécutables par des automates (abstraites). On peut exiger plus et désirer entrer dans un véritable processus d'expérimentation où une machine réelle (ordinateur) est utilisée comme un "simulateur" destiné à produire des textes artificiels que l'on compare ensuite aux textes naturellement produits ou plutôt énonçables par (un ou des) énonciateur(s) humain(s). Cette expérimentation qui impose une programmation de la machine et donc une étude formelle préalable des procédures de traitement des textes, pose des questions théoriques qui ne peuvent rester en suspens : quelles sont les unités taxinomiques ultimes? Y a-t-il des paramètres cachés profonds et moins accessibles à l'observation directe? Quelles sont les opérations et relations qui relient ces entités entre elles? Quel est le statut théorique de ces entités? Quelle est la pertinence théorique des opérations et des relations utilisées? Y a-t-il des procédures algorithmiques qui relient les entités abstraites théoriques aux phénomènes plus directement observables?...

Valider le corps des hypothèses, c'est déduire toutes les conséquences que l'on peut en tirer (en particulier la recherche de la cohérence interne, la confrontation avec d'autres théories et d'autres corps d'hypothèses, un regard sur les domaines voisins qui partagent une partie du thème retenu), puis comparer les objets produits artificiellement et traités par le simulateur avec les objets naturels. Si, aux erreurs d'ajustements, d'approxima-

tions et d'observations évaluées par et dans la théorie, les objets artificiels restent analogues aux objets naturels, alors nous sommes bien en présence d'un véritable processus de validation expérimentale d'hypothèses théoriques.

La linguistique a peu réfléchi à ce que devrait être une expérimentation active. Or l'ordinateur peut jouer ce rôle de synthétiseur (et également d'analyseur automatique): chaque programme informatique destiné à produire "in vitro" un -ou une famille paraphrastique-(d')énoncé(s) est un instrument expérimental déduit de la théorie qui peut être utilisé pour la valider.

1.3. Langues et langage : le problème des catégories

La linguistique a pour objet d'étude "le langage appréhendé à travers la diversité des langues naturelles", en reprenant A. Culioli; elle n'est donc pas "la" science du langage (la philosophie, la psychologie, la logique entre-autres expriment aussi une opinion à propos du langage) mais elle focalise son activité sur les phénomènes langagiers analysables à partir des seules traces textuelles et manifestations linguistiques observables dans des langues réelles. Celles-ci sont -on le sait- diverses, multiples et apparaissent souvent comme irréductibles les unes aux autres. Or, celui qui s'interroge sur le langage est menacé sans cesse par deux dangers : ou bien, il s'enferme dans un glottocentrisme rassurant qui réduit "le langage" à une seule langue (souvent celle de l'observateur ou à une langue issue d'un espace économiquement dominant) : le langage apparaît alors comme dans un miroir qui reflète les propriétés de cette langue, les catégories langagières restent identifiables (souvent à un changement d'étiquettes près) aux seules catégories linguistiques observées; ou bien, il se coupe radicalement des phénomènes observables (qu'il faut apprendre à maîtriser et évaluer) pour spéculer plus librement sur la nature du langage, sans remarquer que les langues proposent souvent, par la diversité des manifestations linguistiques, des inter-

actions imprévisibles entre catégories et phénomènes empiriques a priori inimaginables mais qui sont néanmoins essentiels et, dès que l'on se refuse à mutiler l'étude du langage en ne traitant que les seules propriétés que l'on sait traiter, ces interactions doivent s'intégrer dans la réflexion théorique et trouver leurs descriptions dans les représentations choisies.

Les linguistes, qui apparemment ne s'occupent que des langues (dans la description d'une langue à tradition orale ; dans une spécialisation en linguistique de terrain ou en études contrastives...), se heurtent vite aux catégories linguistiques et aux correspondances lors d'un transfert de langue à langue, ce qui revient finalement à aborder le problème du statut et de la description des catégories langagières. D'une façon générale, la définition précise des catégories en linguistique est souvent escamotée parce que délicate bien que fondamentale. Si l'on se fait une vision assez vaste de la linguistique, on ne peut se contenter de catégories linguistiques assimilées aux "parties traditionnelles" du discours. Les procédures d'analyse distributionnelles établissent quant à elles des classes d'équivalence hiérarchisées, ce qui autorise une certaine réflexion sur chaque langue et même permet d'entreprendre une analyse transformationnelle des phrases (et donc des relations paraphrastiques), sans toutefois bien poser le problème des comparaisons de catégories de langue à langue.

Les catégories linguistiques, obtenues par des procédures taxinomiques, varient selon les langues et ne sont pas des données immédiates d'observation mais sont le point d'aboutissement d'un traitement systématique des données linguistiques ainsi organisées. Que les catégories varient selon les langues, chacun le conçoit et l'accepte sans pourtant en tirer les conséquences méthodologiques qui s'imposent. Souvent, croyant traiter de linguistique générale, on reste enfermé dans un groupe homogène de langues génétiquement apparentées et on présente comme universelle une catégorie (ou une opposition catégorielle) qui est en fait spécifique de ce groupe. Conscient de

la difficulté de passer de catégories spécifiques et particulières à des catégories générales, on change les étiquettes (mais hélas, les étiquettes ne suffisent pas à constituer les catégories) ou encore on transporte une opposition catégorielle (qui fonctionne assez bien dans une langue) dans une autre langue sans toujours s'assurer de la validité de ce transfert. Ainsi, pour fixer les idées, l'opposition spécifique entre "groupe nominal" et "verbe" (catégories morpho-syntaxiques) se transforme biunivoquement et simplement par une opposition "nom" et "prédicat" qui serait plus générale, plus universelle, puisque plus abstraite, (en quoi serait-elle plus universelle s'il y a une bijection canonique entre les deux couples de catégories? où est l'abstraction, sinon que les termes "nom", "prédicat" ont un statut précis en logique mais pas en linguistique?).

Le transfert et la comparaison des catégories d'une langue à une autre pose au descripteur de redoutables problèmes. Pour ne pas prendre des langues éloignées entre elles, donnons un exemple: quelles sont les valeurs des articles postposés aux noms en bulgare et comment "traduire" explicitement ces valeurs en russe (qui ne possède pas d'articles mais un système de cas morphologiques), langue qui est pourtant apparentée au bulgare (groupe slave)? Par quelles règles et valeurs est rendu, en français, le système des articles du bulgare et réciproquement? Y a-t-il des rapports biunivoques? Vraisemblablement pas. De même, si nous prenons l'anglais et le français, l'opposition a (an)/the n'est pas associée terme à terme avec l'opposition un (une)/le (la). La valeur de the n'est pas identique à celle de le (la). Comment établir une correspondance significative entre, d'une part, le système des temps d'une langue qui fonctionne principalement sur des oppositions (morpho-syntaxiques) aspectuelles avec, d'autre part, une langue qui fonctionne surtout avec des oppositions morphologiques de temps ("tense") (nous pensons, par exemple, aux langues sémitiques et romanes). Cela signifie-t-il que la notion de temporalité n'existe pas dans le premier groupe et que celle

d'aspectualité serait absente du second? Assurément non!

Une catégorie grammaticale peut être vue comme une relation entre, d'une part un système de notions grammaticales (temporalité, aspectualité, modalité, diathèse, détermination...) et, d'autre part, un jeu de marqueurs, repérables dans le texte, le plus souvent morpho-syntaxiques, comme le temps, l'aspect, des marqueurs d'assertion (ou d'interrogation, de négation, de souhaitable, de probable, d'obligation pour soi ou pour autrui, d'appréciation), de voix, des déterminants (articles, démonstratifs, quantificateurs)... Les catégories grammaticales sont en fait constitutives des catégories linguistiques d'une langue. Mais ces catégories grammaticales, tout comme les catégories linguistiques, ne sont pas des objets invariants par rapport aux langues et, de plus, elles nécessitent des procédures et des raisonnements argumentatifs qui assurent leurs fondements, car il n'y a pas de relations simples et naturelles -par exemple biunivoques- entre notions grammaticales et jeux des marqueurs, entre les catégories grammaticales et les catégories linguistiques (ces dernières étant en grande partie -nous l'avons dit- le résultat de procédures essentiellement taxinomiques). Chaque marqueur morphosyntaxique renvoie en général à plusieurs notions grammaticales qui sont elles-mêmes reliées entre-elles dans un système complexe à dé-mêler.

Ainsi, face aux langues diverses et multiples, les linguistes ont deux attitudes complémentaires: ou bien, il se résolvent à collectionner des phénomènes irréductibles à toute appréhension par un discours généralisant et ils confectionnent des listes de faits exotiques qui les confortent dans l'idée que chaque langue reste différente et particulière; ou bien, ils tentent de (re)-construire (s'ils existent) les invariants langagiers qui se manifestent par divers procédés linguistiques.

Comment se manifestent ces invariants? D'après ce que nous avons dit, certainement pas par des catégories linguistiques. Si l'on cherche à constituer des catégories généralisables et non plus particulières à chaque langue,

il est indispensable de dépasser le niveau des objets construits par des procédures uniquement taxinomiques et organisées à l'aide de techniques maintenant bien éprouvées (découpage en morphèmes, en phrases grammaticalement bien formées; analyses transformationnelles de phrases complexes en phrases constituantes; analyses syntagmatiques ou équivalentes...) pour mettre au point des procédures d'analyse plus raffinées où entrent en jeu des paramètres métalinguistiques qui ne sont plus biunivoquement associés à des marques observables ou à des classes d'équivalence d'objets directement observables.

Ce dépassement des procédures strictement classificatoires (qui, remarquons-le, ne condamne pas les classifications indispensables à chaque langue) est rendu nécessaire non seulement lorsque l'on recherche les invariants inter-langues mais aussi, lorsque l'on fait éclater le cadre trop rigide de la phrase bien formée grammaticalement pour aborder les problèmes d'énoncés bien constitués et, de ce fait, interprétables.

Les langues constituent la base empirique de la linguistique; cela suppose donc que l'on précise les conditions d'observations (quoi, comment, à partir de quoi observer), c'est-à-dire, qu'il faut se donner et une théorie des observables (problème des données) et une théorie de l'observation, (conditions d'observations, ajustements entre représentations et données) ce qui revient à indiquer quelles sont les procédures explicites d'enregistrement et de traitement des données (systèmes de représentations et de calculs), quelles sont les consignes d'ajustement des objets construits (suites de symboles et de signes) aux objets physiques observables (chaînes sonores ou graphiques), quelles sont les règles d'approximation analogues à un vérifiable calcul d'erreurs (évaluées dans et par la théorie).

L'analyse des rapports langage-langues est alors conceptualisée par une démarche qui s'appuie sur une base empirique (les langues), retourne aux objets le plus directement observables (les énoncés, les familles paraphras-

tiques et paradigmatiques d'énoncés, les textes). Dans l'état actuel des techniques d'observation et d'enregistrement en linguistique (le linguiste observe la plupart du temps à "l'oeil nu" et dispose de bien peu d'appareils capables de traiter les phénomènes de prosodie, de courbe mélodique, de rythme, de pause ...), l'énoncé apparaît comme un objet assez directement observable (du moins pour les problèmes généraux de syntaxe), en tout cas plus directement observable que le phonème, le mot, le syntagme et la phrase. L'énoncé a en fait un double statut, c'est un objet directement observable mais c'est aussi un objet déjà chargé de théorie et produit par la théorie. Mais l'énoncé n'est pas étudié pour lui-même; il appartient à toute une famille paradigmatique et prend ainsi une valeur par rapport aux autres énoncés de cette famille. C'est donc par une étude systématique des déformations (énonciatives, d'une part, et prédicatives, d'autre part) que l'on fait apparaître et l'ambiguïté (un énoncé a souvent plusieurs valeurs) et les relations paraphrastiques et la valeur des différents marqueurs grammaticaux ce qui contribue à construire des catégories grammaticales de plus en plus générales.

2. SYSTEMES DE REPRESENTATIONS ET OPERATIONS METALINGUISTIQUES

Le recours aux paramètres métalinguistiques qui interviennent dans la description des langues -et dont il faut argumenter très sérieusement l'introduction- fait apparaître des systèmes de représentations métalinguistiques (SR μ) que, pour mieux contrôler, on rend externes aux langues naturelles. Ces systèmes engendrent des formules métalinguistiques représentatives d'énoncés; chaque formule étant le résultat d'un agencement formel d'opérateurs et d'opérandes. Lorsque les opérateurs agissent sur leurs opérandes, il s'effectue des opérations qui sont constitutives de la signification attachée à l'énoncé décrit. Chaque enchaînement d'opération impliquées par une formule métalinguistique (conçu comme un véritable programme) construit une signification, résultat de l'effectuation ordonnée des opérations.

Nous nous situons dans la problématique générale de l'énonciation et formulons l'HYPOTHESE GENERALE suivante :

"chaque énoncé (plus généralement texte) est produit (et interprété) par un énonciateur qui met en oeuvre un ensemble d'opérations de production (et d'interprétation ou de compréhension)".

Nous y adjoignons l'HYPOTHESE COMPLEMENTAIRE :

"ce qui varie de langues à langues, ce ne sont pas les opérations mais l'enchaînement de ces opérations impliquées par les formules métalinguistiques représentatives des énoncés. Par ces différences d'enchaînement, on peut expliquer la diversité des catégories linguistiques et la spécificité de chaque langue".

2.1. INVARIANTS METALINGUISTIQUES

Qui effectue l'opération ? Comment sont-elles enchaînées?

La réponse à la première question est simple. Chaque opération renvoie à l'activité métalinguistique du linguiste qui, par ce moyen, représente, décrit, simule, voire explique l'activité de production et d'interprétation d'énoncés par des énonciateurs humains. C'est donc le linguiste qui définit et effectue les opérations. Ce n'est pas un sujet "plein", psychologique, intentionnel, affectif, réel et concret même si le système de représentations contient - en tant que paramètre - un énonciateur. La deuxième question est plus technique. Rappelons qu'une opération est un processus qui fait correspondre un résultat à une opérande. Cette correspondance est établie au moyen d'un opérateur qui est dit "agir sur" ou "porter sur" l'opérande.

L'opérande est constituée d'une séquence ordonnée d'objets qui appartiennent à un domaine; le résultat appartient lui aussi à un codomaine. Une opération est donc complètement définie si : (a) on connaît l'opérateur (ou le processus opératoire de mise en correspondance), (b) le domaine et (c) le codomaine. Chaque domaine (et codomaine) est un ensemble d'objets de même sorte

"nom", "prédictat", "proposition", "valeurs de vérités"...). Il est alors indispensable de préciser les sortes d'objets sur lesquels on opère en particulier celles qui sont primitives (au double sens de premier et de simple) et celles qui sont dérivables des premières. Ces sortes primitives sont des entités formelles qui représentent les catégories métalinguistiques (primitives).

Un système de représentations métalinguistiques (SR μ) est défini par la donnée de :

- (1°) des catégories métalinguistiques (primitives ou dérivées au moyens d'opérations précises) auxquelles on associe des classes (souvent des ensembles) comprenant des objets métalinguistiques qui serviront d'arguments aux opérations; dans certains cas, ces classes seront munies d'une structure algébrique, logique ou topologique selon les catégories;
- (2°) un système d'opérateurs métalinguistique qui induit, dès que l'on précise les domaines et les processus opératoires, un système d'opérations;
- (3°) un réseau (descriptible en partie seulement dans un langage mathématique) de relations entre les objets métalinguistiques, exprimant :

- (a) des relations sémantico-syntaxiques (caractérisation plus ou moins animée, description formelle de processus et état, relations de repérage, localisation spatiale...);
- (b) des relations physico-culturelles liées à une théorie de la perception et des représentations cognitives et culturelles.

La mise en place de tels systèmes est une nécessité dès que l'on veut expliquer 1°) comment sont constituées les catégories grammaticales d'une langue et 2°) comment s'effectue le transfert (non biunivoque) d'une langue à une autre. Une catégorie grammaticale d'une langue (nom, verbe, genre, nombre, détermination, voix, temps, aspect...) est la mise en correspondance entre, d'un côté, un système d'opérateurs (et d'opérations) d'un (SR μ)

et, de l'autre, un ensemble (hiérarchisé) de classes d'équivalence; toute notion grammaticale (temporalité, aspectualité, quantification, détermination, modalité...) est le résultat de diverses opérations ; elle est reliée à un marqueur (contigu ou non) de l'axe syntagmatique, c'est-à-dire, en fait, à un représentant d'une classe d'équivalence. Chaque catégorie grammaticale est donc une relation complexe entre un système de notions grammaticales, sous système d'un $(SR\mu)$ et un jeu de marqueurs morpho-syntaxiques.

La recherche des invariants métalinguistiques consiste, pour nous, à définir les catégories métalinguistiques primitives et les opérateurs élémentaires (dont il faut -rappelons-le- argumenter le statut par rapport à une théorie de l'activité langagière).

Cette hypothèse est différente de celle de S.K. Shaumyan qui suppose l'existence d'un "langage génotype", certes reconstruit mais invariant par rapport aux langues ("Filosofskie Problemy Teoreticeskoj Linguistiki", Moskva, Nauka, 1971). L'hypothèse, à laquelle nous nous attachons, peut être formulée plus techniquement (par rapport à la première formulation adoptée plus haut) de la façon suivante :

"Ce sont les catégories et opérateurs métalinguistiques constituants des $(SR\mu)$ qui restent invariants par rapport aux langues et non les règles de composition et d'agencement. Ces règles, combinées entre-elles, engendrent les diverses formules métalinguistiques de chaque $(SR\mu)$ elles sont ainsi formulées à l'aide des mêmes opérateurs invariants mais les règles qui enchaînent les opérations induites restent spécifiques; en d'autres termes, si $(SR\mu_i)$ est le système de représentation de la langue L_i ($i = 1, 2$), $(SR\mu_1)$ et $(SR\mu_2)$ n'ont aucune raison d'être à priori isomorphes, tout en étant comparables puisque formulées avec les mêmes catégories et opérateurs ce qui justifie la recherche d'éventuels foncteurs de $(SR\mu_1)$ vers $(SR\mu_2)$ ".

Cette hypothèse permet de poser, de façon raisonnable, à la fois et le problème de la diversité des langues et le problème de la traduction et le pro-

blème de l'activité de langage...

Si, comme nous l'avons remarqué, les catégories linguistiques d'une langue ne sont pas directement observables (mais le résultat d'une procédure), les catégories métalinguistiques et les opérateurs le sont encore moins; ils sont reconstruits à la suite de procédures argumentatives contraintes par trois types d'exigences :

(1) réduire (c'est-à-dire, ne pas multiplier, sauf par nécessité) le nombre des catégories métalinguistiques, ainsi que le nombre des opérateurs des (SR μ);

(2) donner un statut théorique (métalinguistique) à chaque occurrence de segment linguistique observable (morphèmes, mots, syntagmes...); donner une réalisation, c'est-à-dire, associer une marque observable (trace) à chaque enchaînement d'opérations métalinguistiques (il s'agit d'un agencement selon la syntaxe de (SR μ));

(3) chaque agencement métalinguistique significatif (ou formule métalinguistique construite par la syntaxe de (SR μ)) doit être glosable dans la langue L (dont (SR μ) est le système de représentation) par un énoncé dit métalinguistique; les gloses de chaque descripteur (qu'il soit linguiste ou simple informateur) sont utilisées en tant que procédure de construction des systèmes de représentations, du moins dans un premier temps. Il s'ensuit que chaque SR μ est générateur d'une métalangue construite et, en même temps, est mis en place à partir des gloses métalinguistiques donc des énoncés d'une métalangue peu contrôlée car plus ou moins spontanée mais néanmoins également objet d'observation pour le linguiste.

Remarquons que le système des gloses (ou énoncés métalinguistiques) est insuffisant pour donner des descriptions entièrement satisfaisantes, cohérentes, complètes et transmissibles, encore moins pour espérer atteindre un stade explicatif, (même si on se donne des règles strictes d'emploi des gloses au sein de la métalangue construite). Les (SR μ) que nous envisageons ne sont pas la traduction symbolique des gloses. Ces (SR μ) sont "extérieurs"

aux langues, ils tendent à éviter les pièges terminologiques et les gloses autoréférentielles ou autodéfinissantes puisqu'ils obligent les linguistes à prendre position : quelles sont les entités primitives, quelles sont les opérations valides, peut-on réduire le nombre des objets primitifs, qu'est-ce qui justifie ce choix?... Enfin, ils permettent des manipulations contrôlées par des procédures explicitables et donc des calculs neutres dégagés du subjectivisme toujours latent du linguiste (du moins à un degré évaluable par chacun et qui peut être ainsi, éventuellement, remis en cause).

Notre objectif est donc :

(1) une recherche d'invariants (catégories et opérateurs) métalinguistiques, (2) qui sont les constituants des (SR_μ) dont (3) il faut rechercher les règles (syntaxe) qui engendrent les formules métalinguistiques significatives, (4) reliées aux énoncés observables par des règles d'ajustement et d'approximation (règles de réalisation).

Le sous-ensemble des formules métalinguistiques significatives de (SR_μ) associés aux énoncés d'une langue L, s'appelle métalangage μL de la langue. L'ensemble des gloses (ou énoncés métalinguistiques) constitue la métalangue de L. Un énoncé E de L est donc décrit ou par une formule métalinguistique significative r(E) intégrée dans un réseau de formules significatives (le système de représentation des paraphrases et déformations grammaticales de E), ou par un ensemble d'énoncés métalinguistiques (gloses) de la métalangue, ou encore par les deux: formules et gloses se renvoyant comme dans un miroir.

2.2 PARAMETRES METALINGUISTIQUES

Quelles sont les catégories métalinguistiques et opérations primitives? Que faut-il entendre exactement par "production" et par "énonciateur"?

De nombreuses descriptions de langues font intervenir de façon souvent non réglée des expressions comme "sujet énonciateur", "moment de l'événement", "point de référence", "situation d'énonciation", "référentiel", "ré-

férence". Les définitions comme "topic / comment" font, par exemple, sans cesse appel à l'un de ces termes : "le locuteur annonce un topique, puis il en dit quelque chose" (Hockett); "le topique ou sujet du discours est défini comme l'élément qui est donné dans la situation générale ou dans une question explicite à laquelle le locuteur répond; le commentaire est défini comme la partie de l'énoncé qui ajoute quelque chose de nouveau (ou qui communique une information à l'auditeur)" (J. Lyons).

Divers articles introduisent, souvent subrepticement, divers paramètres métalinguistiques dans les descriptions. Les paramètres qui sont nécessaires à certaines descriptions sont, selon nous :

- . \mathcal{S}_0 ou origine des énonciateurs ;
- . \mathcal{E}_0 ou origine du repérage aspectuo-temporel ;
- . Sit $(\mathcal{S}_0, \mathcal{E}_0)$ ou origine des situations spatiales ; la situation
- . Sit $(\mathcal{S}_0, \mathcal{E}_0)$ est la situation d'énonciation créée par l'énonciateur à l'instant \mathcal{E}_0 .

Ces paramètres sont les origines des diverses composantes de l'espace énonciatif. On établit un parallélisme entre l'argumentation de E. Benvéniste dans "Appareil formel de l'énonciation" (Langages n° 17, Didier-Larousse, 1973) et la liste des paramètres proposés ici. Ce qui caractérise une telle approche, c'est la théorisation des relations entre énonciateurs, énoncés et relations prédicatives prises en charge par les énonciateurs au travers des énoncés. On ne suppose donc pas un sujet énonciateur universel \mathcal{Y}_u , omnicompétent tant en production qu'en compréhension, témoin objectif d'une langue, capable d'engendrer et de reconnaître toutes les phrases de la langue et elles seulement; si le sujet est universel, il intervient de façon constante dans toute production ou reconnaissance de phrase; il est alors naturel de simuler directement son comportement par un automate universel qui engendre (grammaire formelle) et qui reconnaît (automate accepteur). Il est alors clair que tout système de représentations compatible avec l'hypothèse qui affirme l'existence d'un sujet énonciateur universel, ne représentera pas

l'énonciateur puisqu'il intervient partout de façon constante.

Nous ne critiquons pas ici l'apport fondamental de N. Chomsky à la linguistique contemporaine mais nous désirons centrer notre intérêt non seulement sur la bonne construction des phrases mais aussi sur la bonne constitution des énoncés; ne pas travailler exclusivement sur des descriptions de langues de grandes civilisations mais essayer de tenir compte également des langues à tradition orale ou à diffusion plus restreinte; inclure dans la construction théorique, et cela dès son principe, les rapports dialectiques entre langage et langues; intégrer dans la problématique les opérations de référenciation qui interviennent dans toute production (et interprétation) d'énoncés par des énonciateurs; construire explicitement à l'aide d'opérations générales toutes les catégories linguistiques d'une langue à partir de catégories métalinguistiques plus primitives et généralisables à plusieurs langues.

Nous proposons des descriptions (d'une même langue ou de langues distinctes) où interviennent explicitement les paramètres énonciatifs $\mathcal{I}_0, \mathcal{C}_0$, $\text{Sit}(\mathcal{I}_0, \mathcal{C}_0)$ ainsi que des paramètres dérivés par certains opérateurs précis; leur introduction ouvre une voie pour donner des statuts théoriques à certains "pseudo-concepts", souvent qualifiés de pragmatiques, rencontrés dans la littérature linguistique comme "topic", "comment", "focalisation", "thématisation", "relations casuelles", "énonciation", "acte performatif"... Il est cependant évident que de telles descriptions, formulées dans des systèmes de représentations métalinguistiques (SR_μ), s'appuient, tout en les dépassant, sur des analyses classificatoires précises pour chaque langue. Ces analyses et les systèmes de représentations associés (arbres syntagmatiques, projectifs..., expressions bien parenthésées, structures prédicatives emboîtées (nested predication)...) se présentent comme une première approximation : ce sont souvent, d'assez bonnes descriptions des opérations spécifiquement prédicatives.

2.3 VALEURS REFERENTIELLES ET REFERENTS

Pour résumer, disons que la description d'une langue, du moins telle que nous la concevons, se ramène à la mise en correspondance entre 2 systèmes :

- un système classificatoire (ou linguistique) composé de classes et d'opérations sur les classes (hiérarchies, emboîtements, imbrications, transformations (équivalences ou relations de préordre), entre phrases. Chaque classe est construite par des relations d'équivalence; diverses relations de préordre et d'ordre relient les classes entre-elles*. Les procédures employées pour la construction des classes et relations entre classes sont essentiellement taxinomiques.

- un système d'opérateurs (métalinguistiques), de catégories métalinguistiques, d'objets catégorisés, de domaines, de codomaines, de relations et d'opérations définies sur des catégories (ce qui permet d'engendrer des catégories dérivées issues de catégories plus primitives). Une catégorie métalinguistique n'est pas une simple classe figée obtenue par une simple procédure classificatoire, comme peut l'être une classe distributionnelle, mais est structurée par diverses opérations et mise en place à la suite de procédures argumentatives.

La recherche de l'invariance et l'établissement du généralisable sont fondamentaux dès que l'on vise l'organisation des catégories, de préférence à celle des termes lexicaux. Il serait cependant illusoire -sauf en première approximation- de vouloir séparer complètement, d'un côté, le lexique (conçu comme un stock de racine ou de lexèmes prêts à s'actualiser dans l'occurrence d'une phrase, à condition d'y adjoindre quelques marqueurs morphologiques grammaticaux) et, de l'autre côté, les catégories grammaticales. Une description raisonnée et fine du lexique d'une langue a recours aux catégories grammaticales qui ainsi serviront à organiser quelques grandes classes lexicales. Les problèmes lexicaux ne paraissent cependant pas facilement généralisables

*Voir par exemple Z.S. Harris : Mathematical Structure of Language, Wiley, New-York, 1968.

car trop irréductibles et spécifiques à chaque langue et les procédures d'organisation d'un lexique ne sont guère transférables d'une langue à l'autre.

L'étude des principales propriétés des systèmes de représentations métalinguistiques (SR μ) construits pour les diverses langues naturelles est une étape indispensable pour la conception d'un modèle général de l'activité langagière. En outre, à partir de tels systèmes, on peut montrer comment un sujet énonciateur (qui produit un énoncé) et le sujet co-énonciateur (qui comprend l'énoncé produit) construit tout un réseau de valeurs référentielles qui s'opposent les unes aux autres et participent ainsi à la signification de l'énoncé.

Chaque énoncé E d'une langue donnée L est représenté par une formule métalinguistique $r(E)$ d'un système SR $\mu(L)$ représentatif de la langue L. Cette formule est paramétrée (il intervient nécessairement des occurrences des paramètres $\mathcal{I}_0, \mathcal{C}_0, \text{Sit}(\mathcal{I}_0, \mathcal{C}_0)$), ce qui signifie que la formule $r(E)$ représente la production de E pour un énonciateur \mathcal{I}_0 situé dans un univers temporel ayant pour origine \mathcal{C}_0 et dans un espace organisé à partir de la situation d'origine créé par \mathcal{I}_0 en \mathcal{C}_0

Le paramètre \mathcal{I}_0 ne représente pas un sujet énonciateur "plein" ou concret (donc doué d'une intention psychologique ou psychanalytique); \mathcal{I}_0 peut être considéré comme un sujet énonciateur "vide" compatible avec toute spécification, d'où son caractère général et non contingent. De même, \mathcal{C}_0 est un paramètre à l'origine du système aspectuo-temporel mis en place par l'énonciation: c'est un instant fixe mais compatible avec toute concrétisation temporelle précise. La situation $\text{Sit}(\mathcal{I}_0, \mathcal{C}_0)$ est un domaine spatial abstrait, dépendant de \mathcal{I}_0 et \mathcal{C}_0 compatible avec une délimitation spatiale précise. Les paramètres $\mathcal{I}_0, \mathcal{C}_0, \text{Sit}(\mathcal{I}_0, \mathcal{C}_0)$ ne sont pas astreints à une utilisation par tel ou tel sujet énonciateur concret et "plein", à tel ou tel moment précis du calendrier, dans telle ou telle situation spatiale déterminée, ce sont des paramètres formels (analogues aux paramètres formels des procédures d'un langage de programmation de haut niveau) auxquels il est possible d'affecter des

valeurs précises.

L'usage des paramètres métalinguistiques $\mathcal{I}_0, \mathcal{C}_0 \text{ Sit}(\mathcal{I}_0, \mathcal{C}_0)$ est à l'origine de la distinction valeur référentielle / référent. En effet, une formule métalinguistique paramétrée $F(\mathcal{I}_0, \mathcal{C}_0, \text{Sit}(\mathcal{I}_0, \mathcal{C}_0))$, qui représente un énoncé E de L, s'interprète dans un espace énonciatif; ce dernier constitue le "modèle" du système de représentation métalinguistique SR (L), il est destiné à décrire la signification des énoncés. L'espace énonciatif a une structure multidimensionnelle complexe, descriptible en partie dans un langage géométrique et topologique. Cet espace énonciatif est utilisé pour exprimer les valeurs référentielles des énoncés. Ces valeurs sont relatives aux origines de l'espace énonciatif : l'énonciateur, l'index aspectuo-temporel et la situation énonciative.

Dès que l'on affecte un sujet énonciateur concret, un instant précis, un espace déterminé ou plutôt des représentants de ces domaines, on construit, par l'intermédiaire des valeurs référentielles, les référents associés aux énoncés.

Les structures de l'espace énonciatif (où sont exprimées les valeurs référentielles) et celles de l'espace des référents ne sont pas isomorphes. Ce dernier se trouve muni d'une structure proche de celle qui est sous-jacente à l'espace construit par une grammaire de la perception. Toutes ces questions font partie d'une problématique plus vaste, celle de la référence qui dépasse les seules compétences de la linguistique.

3. HYPOTHESES FONDAMENTALES

Le programme proposé est immense. Il est donc naturel de restreindre, dans une première étape, le domaine des recherches, de définir des objectifs accessibles ainsi qu'une méthodologie destinée à faciliter et rendre contrôlable notre travail.

Toute recherche scientifique se doit de formuler des principes et hypothèses validables. Par définition un principe n'est pas validable, seules

les conséquences que l'on peut en tirer le sont. Il est également possible de formuler des hypothèses compatibles avec les principes et de chercher à valider ce corps des hypothèses. La validation de ces hypothèses n'est souvent pas directe alors que les conséquences que l'on en tire conduisent à formuler des hypothèses locales plus restreintes, plus précises et directement validables (et donc réfutables).

3.1. Subjectivité du langage

PRINCIPE (de la subjectivité du langage en tant que système de représentation) Tout énoncé est produit par un sujet énonciateur. En effectuant son acte d'énonciation, celui-ci utilise un système de représentations autonome et détachable de la réalité extra-linguistique ; ce qui le conduit à construire un référentiel "subjectif" qui suppose une origine locale englobant l'énonciateur, une origine du référentiel temporel et une origine du référentiel spatial.

Ce principe affirme que tout énoncé est produit par quelqu'un qui, lorsqu'il produit son énoncé, organise un référentiel "subjectif", la "subjectivité" se décomposant en trois domaines: les énonciateurs, le temps, l'espace.

De ce principe, on déduit aussitôt le COROLLAIRE Il n'existe pas de sujet énonciateur universel.

Autrement dit, chaque description d'un énoncé, chaque représentation métalinguistique doit laisser une place au sujet énonciateur, à l'origine du temps et l'origine spatiale. Le principe et son corollaire ne sont pas compatibles avec la plupart des modèles qui postulent un sujet omniscient et témoin de la langue, avec la compétence de produire et de reconnaître tout ce qui est un énoncé d'une langue en excluant tout ce qui n'est pas énoncé. Les représentations métalinguistiques compatibles avec le principe formulé ci-dessus inscrivent, dans les écritures, la présence des utilisateurs.

3.2. Paramétrage des formules métalinguistiques

Le principe de la subjectivité affirme autre chose. En produisant un énoncé, le sujet énonciateur utilise un système de représentations. Or, pour toutes les descriptions, le système de représentations doit être général et non contingent; il est nécessaire de recourir à des systèmes symboliques où la présence des utilisateurs va être inscrite sous forme de paramètres. Chaque formule métalinguistique qui représente un énoncé est paramétrée. La nature pragmatique de l'activité langagière qui découle du principe de la subjectivité du langage est traitée dans la représentation métalinguistique elle-même et non pas rejetée dans une composante qui viendrait se surajouter aux composantes syntaxique puis sémantique. Énonciateur, index temporel, domaine ^{du} spatial qualitatif qui traditionnellement relèvent de la pragmatique sont représentés sous forme de paramètres dans un système symbolique destiné à décrire un énoncé.

Ainsi, tout ce qui concerne la personne (je, tu, il), le temps et l'aspect (présent, passé, futur, accompli, inaccompli etc...), les déictiques spatiaux (ici/là-bas/ailleurs), ou temporels (maintenant, aujourd'hui, hier, demain...), les démonstratifs (ce livre,...) etc.... sont traités et analysés dans un système métalinguistique paramétré. (cf. Bibliographie)

Nous pouvons donner une formulation plus technique compatible avec le principe précédent.(cf. §2.2)

HYPOTHESES(*du paramétrage des formules métalinguistiques*): Chaque description d'énoncé fait intervenir une formule métalinguistique représentative de l'énoncé. Cette formule est paramétrée. On retient trois paramètres fondamentaux (non indépendants): \mathcal{I}_0 pour énonciateur; \mathcal{C}_0 pour l'origine du système temporel, $\text{Sit}(\mathcal{I}_0, \mathcal{C}_0)$ pour l'origine des situations (spatiales) énonciatives.

En d'autres termes, un énoncé E est représenté par une formule paramétrée $F(\mathcal{I}_0, \mathcal{C}_0, \text{Sit}(\mathcal{I}_0, \mathcal{C}_0))$. Lorsque l'on instancie les paramètres formels $\mathcal{I}_0, \mathcal{C}_0, \text{Sit}(\mathcal{I}_0, \mathcal{C}_0)$ par des valeurs spécifiques, ce qui revient à affecter

à $\mathcal{F}_0, \mathcal{E}_0$, $\text{Sit}(\mathcal{F}_0, \mathcal{E}_0)$ des valeurs particulières, on obtient une formule non paramétrée. Alors que $F(\mathcal{F}_0, \mathcal{E}_0, \text{Sit}(\mathcal{F}_0, \mathcal{E}_0))$ sera utilisée pour construire le système des valeurs référentielles associées à l'énoncé représenté, la formule instanciée est utilisée pour construire le système des référents associés à l'énoncé. (cf. § 2.3)

3.3. Principe du repérage

PRINCIPE (du repérage) : L'opérateur fondamental de repérage, noté ϵ , est constitutif des systèmes de représentations et catégories métalinguistiques.

Dès que l'on analyse le fonctionnement des copules est/a, le jeu des prépositions, la mise en place des différentes catégories grammaticales (temps, aspects, personnes, détermination, diathèse, modalités...) ou le fonctionnement de la prédication il apparaît nécessaire d'utiliser un opérateur abstrait ϵ qui permet de relier deux objets X et Y en une occurrence de relation: "X ϵ Y": X est considéré comme le terme repéré et Y comme le terme repère.

Il est difficile de donner une argumentation complète pour "justifier" l'hypothèse du repérage, d'où sa formulation sous forme d'un principe. Seules les conséquences techniques et empiriques, et la fécondité des analyses permettent de justifier la plausibilité de ce principe. Le "principe du repérage" permettrait de mieux fonder et de compléter "l'hypothèse localiste" (dont l'origine remonte aux stoïciens) selon laquelle le langage est, dans ses ultimes fondements, une abstraction (réitérable) de l'organisation spatiale centrée sur le sujet énonciateur et filtrée par ses organes perceptifs.

Il est nécessaire de préciser la nature de cet opérateur de repérage ϵ en énumérant ses principales propriétés formelles et axiomatiques puis de montrer comment cet opérateur intervient dans l'organisation et la structuration des principales catégories métalinguistiques: les énonciateurs, le temps, l'espace (et même le domaine notionnel), de façon à expliciter

des foncteurs de transformation entre ces catégories fondamentales. En effet, les observations de nombreuses langues ont montré qu'il y avait des liens étroits (certains ont parlé "d'isomorphismes" ce qui, selon nous est trop fort) entre catégories de la personne, du temps, de l'espace et des notions. Pour illustrer ceci, signalons les oppositions analogiques correspondant aux trois premières catégories : (cf. bibliographie)

DOMAINES

personne	temporel	aspectuel	spatial
je / tu	hier aujourd'hui demain	accompli / inaccompli	ici / Là-bas
il	un jour	aoristique	ailleurs

Il est également bien connu que de nombreuses particules fonctionnent aussi bien dans le domaine temporel, que spatial ou notionnel, par exemple :
venir à trois heures / à Paris / à cheval
tenir jusqu'à cinq heures / jusqu'à Marseille / jusqu'au bout (à épuisement)

Il existe de nombreuses ambiguïtés jouant sur plusieurs domaines. Ainsi se suicider au café renvoie aussi bien au domaine spatial (lieu) que temporel (au moment du) que notionnel (avec du).

La notion du repérage (abstrait) permet donc de relier des catégories entre elles en établissant des relations invariantes et sans doute constitutives de l'organisation langagière, c'est du moins l'hypothèse exprimée par A. Culioli et nous nous attachons à celle-ci en l'érigeant en principe.

L'opérateur de repérage $\underline{\epsilon}$ est une primitive du système de représentations métalinguistiques. Il en a été donné une première description axiomatique*. Cet opérateur reçoit trois spécifications : l'identification, la diffé-

*J.P. Desclés : opérateurs/opérations : méthodes intrinsèques en informatique fondamentale ; applications aux bases de données et à la linguistique, thèse d'état, Université René Descartes, 1980 - voir chapitre IX.

renciation, la mise en rupture (ou "ruption"). Il est constitutif des principales catégories métalinguistique comme cela a déjà été illustré dans de nombreux articles * . Cet opérateur de repérage $\underline{\epsilon}$ est un candidat à l'invariance. Il s'agit là non d'un résultat mais d'une hypothèse de travail qui jusqu'à présent a été féconde.

Ayant admis les deux principes, nous pouvons donner une forme plus technique à l'hypothèse (du paramétrage):

HYPOTHESE : (du paramétrage des formules métalinguistiques : 2ème formulation). Tout énoncé est représenté par une formule métalinguistique paramétrée où une relation prédicative (simple ou complexe) est repérée (de façon directe ou médiatisée) par rapport à une situation énonciative d'origine $Sit(I_0, T_0)$ créée par un énonciateur I_0 et en un instant T_0 , origine de l'organisation du temps linguistique.

L'hypothèse précédente peut alors être résumée à l'aide du schéma de représentation

$$E \xleftarrow{\text{réalisation}} X \in \text{Sit}(I_0, T_0)$$

où X désigne une relation prédicative (simple ou complexe); $Sit(I_0, T_0)$ la situation d'énonciation; I_0 un paramètre appelé "énonciateur"; T_0 un paramètre fixant l'origine temporelle de l'énonciation.

3.4. Invariance des opérateurs

Rappelons une hypothèse déjà présentée et commentée.

"L'acte d'énonciation" effectué par un sujet énonciateur est un programme destiné à produire un énoncé. Ce programme est un enchaînement d'opérations. Le programme, attaché à la formule métalinguistique qui représente l'énoncé indique comment a été constitué l'énoncé produit par un énonciateur. Les opérations, déclanchées lors de l'exécution du programme, sont décrites par un agencement d'opérateurs et d'opérandes.

* voir bibliographie.

HYPOTHESE (de l'invariance des opérateurs constitutifs des catégories métalinguistique) Ce ne sont pas les systèmes de représentations métalinguistiques qui sont invariants mais les opérateurs constitutifs des catégories métalinguistiques.

De cette hypothèse, on déduit immédiatement que les systèmes (SR_{*μ*}) varient selon les langues et selon les différents sujets énonciateurs humains. Cette dernière variation permet de rendre compatible une théorie formelle des représentations métalinguistiques avec les variations socio-linguistiques, tout en séparant ce qui appartient au domaine du généralisable et ce qui nécessiterait des études spécifiques et une approche appropriée. Mais, une théorie des systèmes de représentations métalinguistiques qui vise aussi une approche des phénomènes socio-linguistiques n'est appréhendée que si l'on distingue explicitement le système des "valeurs référentielles" et le système des "référents". (cf. § 2.3)

Cette hypothèse est fondamentale car, si elle est validée, elle donnerait une explication raisonnable à ce qui fait la spécificité de chaque langue, d'une part et à ce qui fait que l'activité langagière se manifeste au moyen d'invariants propres à l'espèce humaine d'autre part. Elle n'affirme donc pas l'existence d'universaux de substance mais seulement des invariants formels et nécessairement abstraits.

3.5. Prédication et énonciation.

Introduisons maintenant une hypothèse plus spécifique.

HYPOTHESE (de la séparabilité -relative- entre prédication et énonciation). Il existe deux ensembles d'opérations: d'un côté les opérations prédictives, d'un autre côté les opérations énonciatives. Ces opérations sont séparables en principe en deux modules disjoints mais un troisième module doit alors nécessairement assembler les résultats des modules de prédication et

d'énonciation.

La prédication concerne essentiellement l'attribution de propriétés à des termes. Cette attribution n'est pas effectuée à l'aide d'une simple opération car le fonctionnement des langues montre que cette attribution s'effectue en plusieurs étapes: choix d'un thème (ou "repère constitutif"), choix d'un premier argument (ou "point de départ" de la relation prédicative), choix d'une orientation de la relation prédicative... bref il ne suffit pas de prendre un prédicat (conçu comme une relation prédicative -ou forme propositionnelle- non saturée) et de constituer une proposition en instanciant par des termes les places non saturées, il faut expliciter comment se constitue la relation prédicative, par quelles opérations élémentaires enchaînées se construit une relation prédicative.

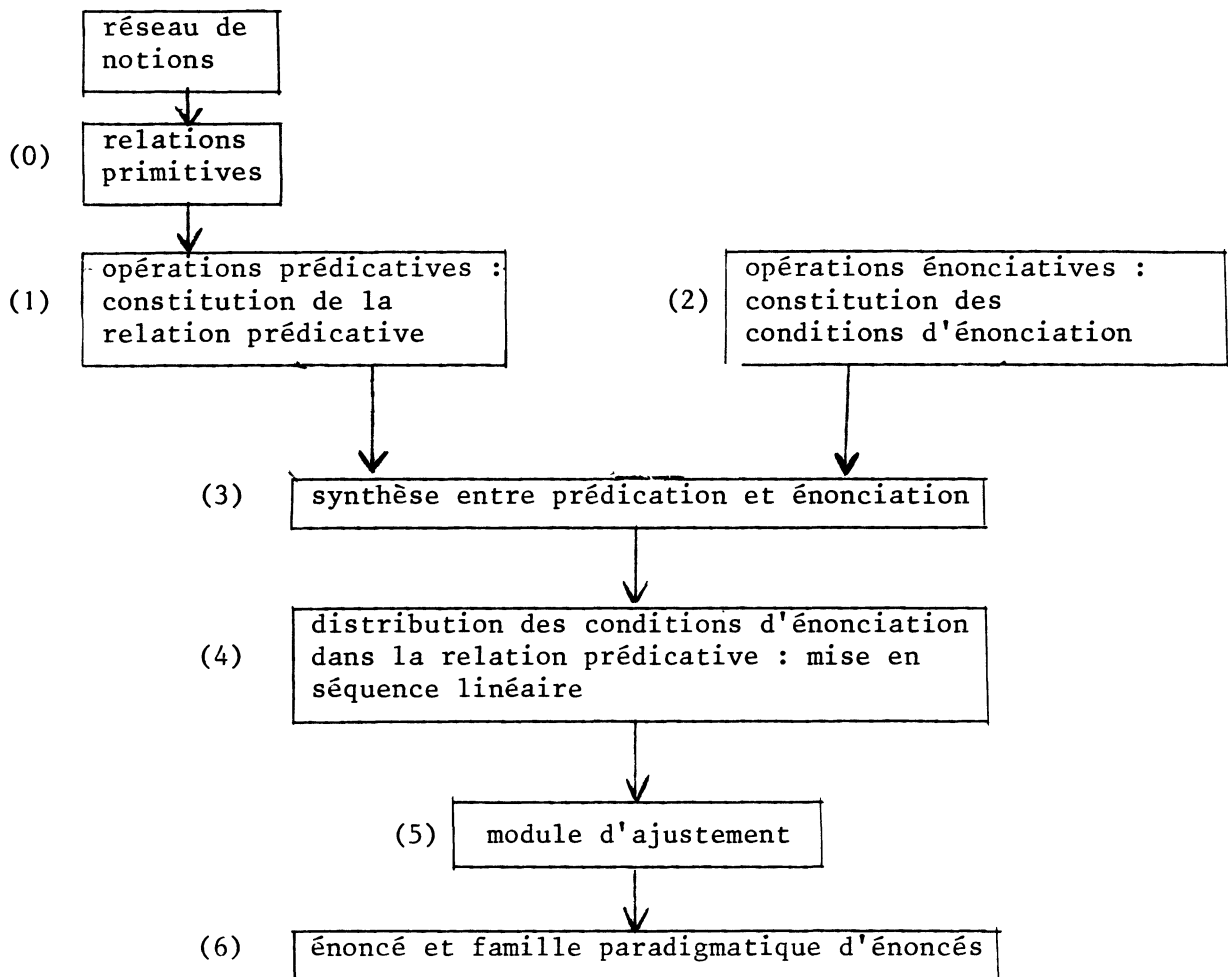
L'énonciation est un ensemble d'opérations élémentaires qui contribuent à fixer les coordonnées énonciatives de la relation prédicative. Celle-ci est repérée dans un espace énonciatif centré sur l'origine complexe $\mathcal{I}_a, \mathcal{I}_0$, $\text{Sit}(\mathcal{I}_0, \mathcal{I}_0)$, au moyen d'occurrences de relations de repérages (par identification, différenciation ou ruption) extrêmement élémentaires. Ce repérage de la relation prédicative contribue à fixer les "conditions d'énonciation" et donc à déterminer les valeurs référentielles de l'énoncé. Dès que l'on déforme les conditions d'énonciation d'un énoncé E, on obtient un autre énoncé E' équivalent à E modulo la relation prédicative commune aux deux énoncés mais évidemment non paraphrastique à E. Les conditions d'énonciations servent à représenter les marqueurs grammaticaux comme les temps, les aspects, certains déterminants, les personnes, les modalités... Le changement des conditions d'énonciation se traduit par un changement des formes grammaticales.

La distinction prédication/énonciation est plus générale que l'hypothèse opposant un modus à un dictum (cette hypothèse remonte au moins aux stoïciens et aux médiévaux), reprise avec des modulations différentes par

G. Frege, C. Bailly, E. Benveniste ou même récemment par Z.S. Harris. Elle diffère dans sa formulation et par la technique qu'elle implique de "l'hypothèse performative" exprimée par la grammaire générative transformationnelle. Le dictum concerne tout ce qui est "disible" (nous dirions tout ce qui est prédiqué) alors que le modus introduit une relation entre ce qui est dit et celui qui le dit (ce sont, pour nous, les conditions d'énonciation).

Dans la relation prédicative constituée, il apparaît, pour certains énoncés, des traces de la situation énonciative (par exemple dans l'analyse de il y a) ce qui montre que la prédication doit être reliée à l'énonciation. Par exemple: Jean est ici est commutable avec Jean est à Paris mais l'analyse de ici, inséré dans la relation prédicative, n'est possible que si l'on se réfère à la situation énonciative d'origine puisque ici renvoie obligatoirement à un lieu où se trouve "celui qui parle", c'est-à-dire l'énonciateur \mathcal{J}_o .

Ceci nous amène à proposer le modèle (en production) des opérations prédicatives et énonciatives. En génération, le module énonciatif est séparé du module prédicatif mais un module synthétise les conditions d'énonciation au résultat de la prédication puis un autre module distribue les valeurs déterminées par les conditions d'énonciation sur les différents termes de la relations prédicative (par exemple, en français, le temps et l'aspect sont généralement attachés morphologiquement au verbe). Un cinquième module assure l'ajustement entre représentation métalinguistique et réalisation attestable (sous forme d'un "énoncé artificiel" engendré, en production, par une machine et de ce fait comparable à un énoncé naturellement produit). A partir de ces modules, il est possible d'engendrer une famille paradigmatique d'énoncés. L'entrée du module prédicatif est un ensemble de relations primitives extraites d'un réseau de notions (ou "prédicables"). Nous avons donc le schéma :



Ce qui précède est un corps de principes et hypothèses sur lequel repose "la théorie des invariants langagiers" élaborée par l'équipe (de A. Culicoli) à laquelle nous sommes attachés. D'autres hypothèses concernant les catégories grammaticales, la constitution de la relation prédictive et de la "lexis", la place et la représentation du lexique interviendraient. Elles prendraient une forme plus technique et feraient appel pour leur formulation à un langage et à des concepts mathématiques (langage de la topologie générale, logique combinatoire et λ -calcul, logique intensionnelle, théorie des préfaisceaux et théorie des topoi...). Nous détaillerons ces hypothèses plus techniques dans d'autres articles.

4. PROCESSUS DE VALIDATION D'HYPOTHESES

Ces hypothèses permettent de construire une théorie qui engendre et justifie les divers systèmes de représentations métalinguistiques (SR μ); si on donne des axiomes (ou schémas d'axiomes) et des règles (ou schémas de règles) de déduction, on calcule toutes les formules métalinguistiques déductibles d'un même schéma canonique (ce qui revient à construire de véritables systèmes formels); on aborde par cette démarche et de façon systématique, le problème de la paraphrase et des déformations énonciatives (étant donné un énoncé, trouver tous les énoncés équivalents déductibles d'un schéma canonique).

La construction de systèmes formels (S.F.) rend envisageable, et seulement à ce stade :

- (1) la vérification de la cohérence de la théorie (est-elle consistante ou non-contradictoire? Est-elle complète? Est-elle stable et indépendante des utilisateurs? Peut-on reproduire les raisonnements et les calculs?);
- (2) la recherche de structures et d'isomorphismes (nous avons déjà signalé que les catégories métalinguistiques étaient structurées);
- (3) la formulation d'éventuelles prédictions sur des faits non encore enregistrés, ce qui est un des moyens de valider la construction théorique;
- (4) le traitement systématique des données sous une forme opératoire qui ne soit pas livrée à l'arbitraire spéculatif ou subjectif du linguiste qui manipule;
- (5) la confrontation de la théorie linguistique ainsi construite avec d'autres théories linguistiques; l'évaluation des équivalences (plus ou moins fortes) ou des puissances relatives entre théories;
- (6) la recherche d'explications générales et l'organisation globale du champ théorique et du champ des phénomènes soumis à une observation plus directe.

La recherche de systèmes formels locaux mais compatibles entre eux,

aboutit à l'élaboration d'un modèle global à vocation explicative. Ce modèle doit se confronter à lui-même (confrontation interne) par la recherche indispensable de cohérence; il se confronte aux autres modèles (confrontation externe) par la recherche d'équivalences et de dépassements; il se confronte aux phénomènes plus directement observables par une série de montages expérimentaux.

La confrontation avec les autres modèles théoriques est facilitée (est-elle possible autrement?) par la mathématisation (comment comparer deux théories écrites dans des métalangues différentes?). Il est alors possible de rechercher d'éventuels morphismes -voire isomorphismes- entre deux systèmes formels.

La cohérence formelle de la théorie est-elle suffisante ? Les systèmes formels (S.F) sont construits pour formaliser une théorie "intuitive" du langage et des langues (vouloir formaliser le langage ou les langues n'a aucun sens !). Une théorie "intuitive" T n'est pas, en général, un objet mathématique. L'adéquation entre une théorie T "intuitive" et un ensemble M de systèmes formels (S.F) locaux n'est donc pas démontrable.

On dit qu'il y a adéquation entre T et M, si (a) toute proposition théorique T est formalisée par un (ou plusieurs) théorème(s) de M (on formule alors des épi-théorèmes) ; (b) tout théorème démontré de M reçoit une interprétation intuitive pertinente dans la théorie intuitive. Le test d'adéquation entre T et M n'est donc pas objet de démonstration mais est le résultat d'un certain consensus plus ou moins large. La formalisation - on le sait - n'est pas neutre : elle impose des décisions et des prises de position ; l'interprétation de son côté doit donner, dans le meilleur des cas, un statut théorique à toutes les entités formelles construites et engendrées par les calculs.

Si le consensus était total, on aboutirait alors à une théorie linguistique formalisée où intuitions linguistique et formelle se traduiraient natu-

rellement l'une dans l'autre. Dans la recherche pratique et quotidienne, une "théorie intuitive" n'est pas -ou rarement- un corps complet -et toujours cohérent- de principes, de propositions théoriques, de procédures argumentatives et déductives qui relient naturellement toutes les propositions empiriques déduites des principes initiaux aux phénomènes accessibles à la simple observation. Il est alors possible et souhaitable -et c'est un des objets principaux de la formalisation- que le (ou les) système(s) formel(s), non seulement aide(nt) à vérifier la cohérence interne de "la théorie intuitive" mais également soi(en)t utilisé(s) pour l'enrichir et la réarranger, le formel participant à l'explication et à la construction théorique.

C'est donc par un balancement continu entre formalisation et interprétation que la théorisation s'organise. C'est dans ce sens que la mathématique est constitutive de la théorie linguistique.

La validation d'une théorie linguistique est envisageable de deux façons complémentaires :

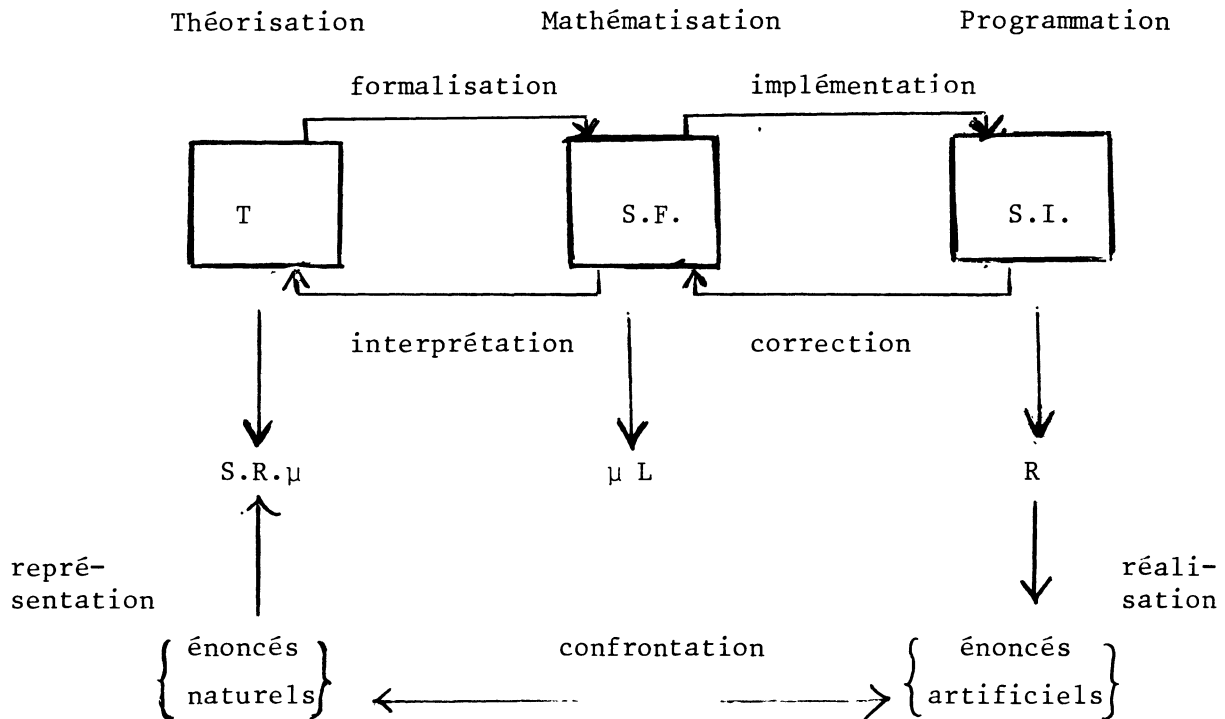
- ou bien par des observations minutieuses sur plusieurs langues mettant en jeu des linguistes (théoriciens et de terrain) et des informateurs;
- ou bien par des expérimentations consistant à construire des appareils déduits de la théorie et capables de reproduire "in vitro" des énoncés.

La construction des systèmes formels (S.F.) locaux qui formalisent adéquatement (à un certain consensus près) une portion de la théorie intuitive T permet d'en déduire des instruments d'expérimentation: ce sont les programmes implémentés (S.I.) qui simulent les systèmes formels (S.F.), ce sont eux les véritables instruments produits par la théorie. Nous en déduisons des protocoles expérimentaux qui confrontent (aux erreurs d'ajustements et d'approximation près, prévus et évalués dans et par la théorie) les énoncés artificiels (produits par la machine) aux énoncés naturels susceptibles d'être produits en situation signifiante par un énonciateur humain.

Nous avons donc trois étapes (voir figure) :

(1) théorisation, (2) formalisation ou mathématisation (3) implémentation ou programmation ce qui représente un processus général de validation des hypothèses théoriques.

Figure :



T = théorie

S.F. = systèmes formels

S.I. = systèmes implémentés

S.R.μ. = systèmes de représentations métalinguistiques

μ L = métalangage; $\mu L \subset S.R.\mu$

R = séquences de symboles résultats de calculs automatisés (programmes).

Il est clair, cependant, que ces mouvements ne se réalisent pas dans un ordre linéaire mais sont organisés selon un processus récurrent et en spirale qui fait apparaître tour à tour: raisonnements et concepts linguistiques, structures et opérations formelles, fonctions d'accès et procédures programmées. On implémente un système et on cherche à savoir ensuite s'il est correct. On formalise une théorie et l'on cherche ensuite à interpréter les

théorèmes démontrés.

5. RECHERCHE DE STRUCTURES

Il est évident que nous ne pouvons pas tout aborder en même temps et qu'il est nécessaire de se doter d'une véritable stratégie.

1. La méthode employée consiste à travailler sur une famille \mathcal{F}_1 d'énoncés (regroupés parce que faisant partie d'un même problème) qu'il faut décrire et analyser en expliquant :

(a) les différences spécifiques : pourquoi telle chaîne diffère de telle autre? à la suite de quelles opérations?

(b) pourquoi certains énoncés sont reconnus "équivalents" (à quelques nuances près qu'il faut préciser) (paraphrases), pourquoi des énoncés très proches du point de vue de leur organisation observable ont des valeurs référentielles distinctes? pourquoi certains énoncés ont plusieurs valeurs (ambiguïté)?

Pour cela, on essaie de donner une représentation métalinguistique à chaque énoncé de la famille (par la méthode des gloses génératrices de représentations formelles) et l'on compare ensuite les formules obtenues.

2. Dans une deuxième étape, on élargit la famille \mathcal{F}_1 , en tentant de :

- rechercher un schéma canonique commun à partir duquel toutes les formules métalinguistiques, associées aux énoncés "équivalents" de \mathcal{F}_1 seront dérivées par des opérations de production (déduction);

- constituer ainsi une famille structurée de formules métalinguistiques qui, pour les unes se réalisent (aux ajustements et approximations près) par des énoncés, pour d'autres par des énonçables, pour d'autres (si le système des règles qui organise les opérations de production n'est pas assez filtrant, donc inadéquat) par des chaînes ininterprétables quelque soient les contextes (donc inénonçables par tout énonciateur).

La famille des énoncés et énonçables ainsi construite constitue alors

une nouvelle famille \mathcal{H}_2 .

3. Arrivé à ce stade, il nous faut :

(c) expliquer pourquoi certaines chaînes sont plus acceptables que d'autres (en se donnant une théorie de l'acceptabilité et une théorie de l'observation); pourquoi certaines configurations de symboles conduisent à des chaînes ininterprétables (incompatibilité des catégories grammaticales);

(d) montrer que les dérivations et déductions effectuées à l'aide des opérations de production conduisent à des familles structurées, d'énoncés et d'énonçables, telles que les relations entre énoncés, la nature des opérations mises en jeu, le statut du schéma canonique commun soient pertinents et interprétables dans la théorie "intuitive".

(e) généraliser le système de représentation et le fragment de système formel dans deux directions: en le rendant compatible à la fois avec d'autres phénomènes ou problèmes linguistiques et également avec les systèmes de représentations d'autres langues analysées pour le même problème.

4. La généralisation (à d'autres problèmes et à d'autres langues) constitue une troisième famille \mathcal{H}_3 d'énoncés englobant nécessairement, par construction, \mathcal{H}_2 et une partie de \mathcal{H}_1 , soit \mathcal{H}_1' . Il est alors possible de recommencer le processus à partir de \mathcal{H}_3 .

La description de cette méthodologie montre que le point de départ est approximatif (sur quels critères constitue-t-on \mathcal{H}_1 ?).

Les étapes ne s'enchaînent pas selon un ordre linéaire mais selon un processus récurrent.

Les systèmes de représentations métalinguistiques (SRμ) sont constitués -nous l'avons dit- à partir des gloses fournies par le linguiste (ou parfois par l'informateur) mais ces gloses restent évidemment insuffisantes car elles ne garantissent nullement la cohérence. D'autre part, les pièges

terminologiques persistent, les ambiguïtés restent délicates... En même temps que le système des gloses, on a recours à des systèmes de représentations "extérieurs" qui font apparaître des régularités structurelles pour, d'une part, construire, et d'autre part, étudier et manipuler les systèmes.

Chaque système de représentations (SR_μ) fait apparaître une (ou plusieurs) classe(s) de structures. Les grammaires "non-contextuelles" de N. Chomsky sont fondamentalement liées aux structures arborescentes. De façon moins évidente, la plupart des théories linguistiques actuelles utilisent abondamment la structure arborescente. On peut argumenter ce fait, sans toutefois l'expliquer autrement que par les considérations sociologiques et techniques suivantes :

- (1°) Les procédures essentiellement classificatoires sont employées pour la plupart des théories linguistiques sans autres procédures complémentaires;
- (2°) la simplicité (peut-être apparente : qu'est-ce qui est simple?) des arborescences et des propriétés récursives qui y sont attachées est transportée dans l'analyse des langues naturelles;
- (3°) l'utilisation inconsciente du modèle des langages de programmation actuels où les structures syntaxiques des programmes sont représentables, en grande partie, en termes arborescents;
- (4°) la connaissance bien établie des propriétés mathématiques et des facilités de traitement: les fonctions d'accès, les implémentations, la gestion et la transduction d'arborescences -ainsi que les algorithmes associés- sont bien connus et concrètement réalisés par les informaticiens;
- (5°) l'influence inconsciente des procédures informatiques de traitement des listes et des langages associés (LISP...) qui risquent, tout en les servant, de conditionner et d'enfermer les théories linguistiques dans un cadre structurel trop restreint et parfois mutilant.

Bien des théories linguistiques se donnent une théorie assez lâche des représentations. Par exemple, le modèle transformationnel "standard" de N. Chomsky représente toute phrase P non par une seule "structure (classifica-

toire) de surface" I_0 (arborescence ou indicateur syntagmatique) mais par une suite ordonnée I_0, I_1, \dots, I_n où I_0 est "la structure (classificatoire) arborescente dite profonde" et I_1, \dots, I_n , sont des structures intermédiaires (également des arborescences). Ainsi, dans ce cadre théorique, aucune des structures ne représente globalement toutes les relations qui apparaissent dans l'analyse syntaxique d'une phrase P; il semble bien qu'aucune structure arborescente ne peut réaliser cette contrainte structurelle.

Notre approche consiste alors à rechercher et à étudier une structure plus générale \mathcal{S} sous-jacente aux systèmes de représentations (SR_μ) (que nous construisons) telle que cette structure soit soumise à cinq ordres de contraintes :

- (1°) étant donné un énoncé E, soit $r(E)$ la formule métalinguistique qui le représente dans (SR_μ); la structure sous-jacente à la constitution de $r(E)$ doit exprimer toutes les relations qui apparaissent dans l'analyse de E et toutes les opérations constitutives de $r(E)$ (et donc E);
- (2°) soient E et $r(E)$; si un jeu d'opérations métalinguistiques de "déformation" de $r(E)$ fait passer de $r(E) = r$ à r_1, \dots, r_n alors les énoncés E, E_1, \dots, E_n associés à r, r_1, \dots, r_n sont les "déformés de E", c'est-à-dire que: $r = r(E), \dots, r_n = r(E_n)$ (problème des paraphrases et déformations);
- (3°) si E est construit, à partir de schémas canoniques, par divers états intermédiaires, alors la représentation $r(E)$ de E doit avoir une structure qui permet de reconstituer les états intermédiaires constitutifs -au moins les plus fondamentaux- de $r(E)$ (donc de E),, c'est-à-dire, en fait, son "histoire constitutive";
- (4°) la structure \mathcal{S} sous-jacente aux formules métalinguistiques doit, lorsqu'on se restreint aux relations purement classificatoires, se dégénérer en structure arborescente;
- (5°) la structure \mathcal{S} doit être compatible (par des procédures algorithmiques

explicites) avec un ordre séquentiel (ou linéaire) propre à l'ordre syntagmatique totalement ordonné de chaque énoncé.

Remarquons que la structure arborescente est directement compatible avec l'ordre linéaire (5°) mais que, si elle représente adéquatement "l'histoire dérivationnelle" d'une phrase dans le cadre syntagmatique, elle ne présente pas son "histoire constitutive" reliant une "structure profonde" à une "structure de surface" dans le cadre des grammaires transformationnelles de Chomsky. Les contraintes imposées à \mathcal{S} sont maximales. La recherche entreprise actuellement consiste à étudier une structure plus générale que l'arborescence et qui appartienne à une classe particulière de graphes. Nous avons appelé "treillage" cette structure et étudions non seulement ses propriétés algébriques mais aussi ses propriétés combinatoires et ses diverses représentations en machine (fonctions d'accès notamment et transformations dynamiques). Il est important de poser ces contraintes pour mieux faire apparaître les structures sous-jacentes à l'organisation même des énoncés car un énoncé n'apparaît pas comme un représentant isolé et unique, mais dans un réseau d'énoncés où il ne prend sa valeur que par rapport aux autres.

BIBLIOGRAPHIE

I. PUBLICATIONS DE L'EQUIPE PITFALL

- [BADE-72] BARBAULT M.C., & DESCLES J.P., Transformations formelles et théories linguistiques, Paris, Dunod, 1972.
- [BCD-76] BESTOUGEFF H., CULIOLI A., DESCLES J.P., "Programme interdisciplinaire de traitement formel et automatique des langues et du langage", Paris, Univ. de Paris 7, Département de Linguistique et Département d'Informatique Générale, (contrat D.G.R.S.T. n°74.70341), Rapport final pour la D.G.R.S.T., Juin 1976.
- [BEDE-74] BESTOUGEFF H., & DESCLES J.P., "L'informatique comme moyen d'expérimentation en linguistique", Analyse et Validation dans l'étude des données textuelles, Ed du C.N.R.S., Paris, 1974.
- [BEDE-76] BESTOUGEFF H., & DESCLES J.P., "Généralisation algébrique du modèle de E.F. Codd", Panorama de la nouveauté informatique en France, t.II, Congrès AFCET, 1976.
- [BEDE-78] BESTOUGEFF H., & DESCLES J.P., "Pitfall", Natural language research program . Natural language communication with computers Novossibirsk, 1978.
- [BLBO-79] BLOSSEVILLE M.J., & BOUILLOUX R., "Conception d'une base de données relationnelle à partir d'une représentation formelle d'un texte en langue naturelle", Participation au colloque de base de données, Univ. de Paris 6, 1979.
- [BLBO-80] BLOSSEVILLE M.J., & BOUILLOUX R., Méthodologie de construction d'une base de données à partir d'un texte la décrivant, Thèse de 3^{ème} cycle, Univ. de Paris 7, Juin 1980.

- [COL-79] COLAS F., Approche théorique de la modalité en linguistique. La modalité appréciative, Thèse de 3^{ème} cycle, Univ. de Paris 7, octobre 1979.
- [COM-77] COMMIOT B., "Algorithmes de manipulations des structures associées aux treilles", Rapport D.G.R.S.T. n°74.70341, Janvier 1977.
- [COR-80] CORI M., Structures hiérarchiques et opérateurs typés : Algorithmes de transformations, Thèse de 3^{ème} cycle, Univ. de Paris 7, 1980.
- [COR-82b] CORI M., "Description d'une classe de grammaires de graphes sans circuit", R.A.I.R.O.-Informatique théorique, vol.16, n°1, 1982.
- [CULI-68] CULIOLI A., "La formalisation linguistique", Cahiers pour l'Analyse, 9, Paris, Seuil, 1968.
- [CULI-71] CULIOLI A., "A propos d'opérations intervenant dans le traitement formel des langues naturelles", Mathématiques et Sciences Humaines, 34, 1971.
- [CULI-79] CULIOLI A., "Valeurs modales et opérations énonciatives", Modèles linguistiques, 1.2, 1979.
- [CULI-80] CULIOLI A., "Valeurs aspectuelles et opérations énonciatives : l'aoristique", La notion d'aspect, Paris, Klincksieck, 1980.
- [CULI-81] CULIOLI A., "Sur le concept de notion", Colloque sur : Détermination, Aspects, Modalités, B.U.L.A.G. 8, Besançon, 1981.

- [CUDE-75] CULIOLI A., & DESCLES J.P., "Considérations sur un programme de traitement automatique des langues et du langage", Informatique et Sciences Humaines, Colloque du C.N.R.S., Marseille, 1975.
- [CUDE-81] CULIOLI A., & DESCLES J.P., "Systèmes de représentations linguistiques et métalinguistiques. Les catégories grammaticales et le problème de la description de langues peu étudiées", Rapport U.N.E.S.C.O., Paris 7, 1981 .
- [DES-74a] DESCLES J.P., "Quelques opérations de prédication", Rapport Pitfall, Univ. de Paris 7, 1974.
- [DES-74b] DESCLES J.P., "Opérateurs, opérations", Rapport technique Pitfall, 7, (contrat D.G.R.S.T. n°74.7.0341), Univ. de Paris 7, Décembre 1974.
- [DES-75] DESCLES J.P., "Systèmes transformationnels de Z. S. Harris : quelques réflexions : I. Construction du noyau ; II. Eclatement du noyau", T-A Informations, 1 et 2, Paris, Klincksieck, 1975.
- [DES-76a] DESCLES J.P., "Description de quelques opérations énonciatives", Modèles logiques et niveaux d'analyse linguistique, Paris, Klincksieck, 1976.
- [DES-76b] DESCLES J.P., "Opérations sur les opérateurs typés", Notes aux comptes rendus de l'Académie des Sciences, T. 283, Série A, 1976.
- [DES-76c] DESCLES J.P., "Transformations d'opérations et de multi-opérations", Notes aux comptes rendus de l'Académie des Sciences, T. 283, Série A, 1976.
- [DES-78] DESCLES J.P., "Enoncés et énonçables", Lingua e stile, Anno XIII, Società Editrice di Mulino, Bologne, Juin 1978.

- [DES-79a] DESCLES J.P., "Représentation formelle de quelques déictiques français", Linguaggi e Formalizzazioni, Rome, Bulzoni, 1979.
- [DES-79b] DESCLES J.P., "Présentation du programme interdisciplinaire de traitement formel et automatique des langues et du langage (Pitfall)", Linguaggi e formalizzazioni, Roma, Bulzoni, 1979, pp.463-489.
- [DES-80a] DESCLES J.P., "Construction de la catégorie grammaticale de l'aspect", La notion d'aspect, Paris, Klincksieck, 1980.
- [DES-80b] DESCLES J.P., "Mathématisation des concepts linguistiques", Modèles linguistiques, T.II,1, Paris, 1980.
- [DES-80c] DESCLES J.P., Opérateurs/opérations : Méthodes intrinsèques en informatique fondamentale. Applications aux bases de données et à la linguistique, Thèse d'état, Univ. René Descartes, Juin 1980.
- [DES-80d] DESCLES J.P., "Recherches sur les opérations constitutives du langage", Fundamenta scientiae 95, Univ. Louis Pasteur, Strasbourg, 1980.
- [DES-82a] DESCLES J.P., "Opérations constitutives d'énoncés", Coll. int. Mathématiques et Linguistiques, Univ. de Neuchâtel, Juin 1980, à paraître sous les actes en 1982.
- [DES-82b] DESCLES J.P., "Quelques réflexions sur les rapports mathématiques et linguistique", livre collectif : Penser les mathématiques, Paris, Seuil, Coll. dirigée par J.M. Lévy-Leblond, 1982.

- [DES-82c] DESCLES J.P., "Quelques systèmes de représentations linguistiques et métalinguistiques", Actes du Colloque sur la notion de système, Lyon, 1980, Aix-en-Provence, Librairie de l'Université, 1982.
- [DEGU-81] DESCLES J.P., & GUENTCHEVA Z., "A propos de la valeur fondamentale du parfait en bulgare Colloque franco-bulgare, Sofia, 1980, Sapistavitelno ezikoznanie, 4, 1981.
- [GUE-76a] GUENTCHEVA Z., "Métalangue, Métalangage, Métalinguistique", Centro Internazionale di Semiotica e di linguistica, Working papers, 60/61. A, Univ. di Urbino, 1976.
- [GUE-76b] GUENTCHEVA Z., Présentation critique du modèle applicatif de S.K. Shaumyan, Paris, Dunod, 1976.
- [KOU-82] KOULOUGHLI D., Sur la phrase nominale en arabe classique. Contribution à l'étude formelle de la catégorie de "mubtada?", (Problèmes de thématization), (Pitfall, 43, 1977) , à paraître dans T.A. Informations, 1, 1982.
- [KUN-81] KUNTZ M., "Dialogues, déictiques et citations", T.A. Informations, 2, Paris, Klincksieck, 1981.

II. REFERENCES BIBLIOGRAPHIQUES.

- [ALL-77] ALLWOOD J., & ANDERSSON L.G., & DAHL O., Logic in linguistics, Cambridge Univ. Press, Cambridge, 1977.
- [BAR-64] BAR-HILLEL Y., Language and Information, Selected essays on their theory and application, Addison Wesley Pub. Co. 1964.
- [BEN-68] BENABOU J., Structure algébrique dans les catégories, Cahiers de topologie et géométrie différentielle, Dunod, Paris, 1968.
- [CHO-68] CHOMSKY N., L'Etude formelle des langues naturelles, Mouton Gauthier-Villard, Paris, La Haye, 1968.
- [CHSC-68] CHOMSKY N., & SCHUTZENBERGER M.P., The algebraic theory of context-free languages, Computer programming and formal systems, North Holland Pub. Co., 1963, en français dans Langages, 9, 1968.
- [CRE-73] CRESSWELL M.J., Logics and languages, Methuen, 1973.
- [CUFE-68] CURRY H.B., & FEYS R., Combinatory logic, North Holland Pub. Co., Amsterdam, 1968.
- [DAMA-76] DAVID J., & MARTIN R., (editors), Modèles logiques et niveaux d'analyse linguistique, Actes du colloque organisé par le Centre d'Analyse Syntaxique de l'Université de Metz, 1974, Klincksieck, 1976.
- [FIT-74] FITCH F.B., Elements of combinatory logic, Yale University Press, 1974.

- [FRE-69] FREGE G., Les fondements de l'arithmétique, Recherche Logico-mathématique sur le concept de nombre, Seuil, Paris 1969, Traduction de C. Imbert de Grunlagen der Arithmetik, 1884.
- [FRE-67] FREGE G., The basic laws of arithmetic, Exposition of the system, Translated and edited with an introduction by Montgomery Furth, Univ. of California Press, Berkley and Los Angeles, 1967, Traduction de Grungesetze der Arithmetik, 1893.
- [GUE-76] GUENTCHEVA Z., Présentation critique du Modèle Applicationnel de S.K. Shaumyan, Doc. Centre Ling. Quantitative, 30, Dunod, Paris, 1976.
- [GRLE-70] GROSS M., & LENTIN A., Notions sur les grammaires formelles Paris, Gauthier-Villard, coll. Programmation, 1970.
- [HAR-68] HARRIS Z.S., Structures mathématiques du langage, Paris, Dunod, 1968.
- [JAK-61] JAKOBSON R., (editors) Structure of language and its mathematical aspects, Proceedings of symposia in Applied Mathematics, Vol. XII, American Math. society, 1961.
- [LAW-63] LAWERE F.W., "Functorial semantics of algebraic theories", P.N.A.S., Vol. 50, 5, 1963.
- [MEL-74] MEL'CUK I.A., Model Smysl-tekst (Modèle "sens-texte"), Moscou, 1974.
- [MON-74] MONTAGUE R., Formal philosophy - Selected papers of Richard Montague, Yale Univ. Press, ed. Richmond H. Thomason, 1974.
- [PAR-76] PARTEE B.H., Montague grammar, Acad. Press, inc. 1976.

- [PEI-34] PEIRCE S., Collected papers, T. I,VI, Harvard Univ. Press, Cambridge, Mass. 1934.
- [SCH-59] SCHÜTZENBERGER M.P., Un problème de la théorie des automates, Séminaire Dubreuil-Piscot, Paris, déc. 1959.
- [SHA-77] SHAUMYAN S.K., Applicational grammar as a semantic theory of natural language, Edinburgh Univ. Press, 1977.
- [STO-77] STOY J.E., Denotational semantics : the scott-strachey approach to programming language theory , the MIT . Press, 1977.

*

*

*

Pour les termes linguistiques, voir par exemple :

POTTIER B., Comprendre la linguistique, Marabout Université, 1975.